Deconvolution of a Distribution Function

Author(s): Clifford B. Cordy and David R. Thomas

Source: *Journal of the American Statistical Association*, Dec., 1997, Vol. 92, No. 440 (Dec., 1997), pp. 1459-1465

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2965416

Deconvolution of a Distribution Function

Clifford B. CORDY and David R. THOMAS

We consider the estimation of a distribution function when observations from this distribution are contaminated by measurement error. The unknown distribution is modeled as a mixture of a finite number of known distributions. Model parameters can be estimated and confidence intervals constructed using well-known likelihood theory. We show that it is also possible to apply this approach to estimation of a unimodal distribution. An application is presented using data from a dietary survey. Simulation results are given to indicate the performance of the estimators and the confidence interval procedures.

KEY WORDS: Expectation-maximization algorithm; Mixture of distributions; Profile likelihood; Unimodal distribution.

1. INTRODUCTION

Our problem is the estimation of the distribution function, F_U, of a random variable U based on observations of random variables of the form $Y = U + \varepsilon$, where ε represents measurement error. For example, the value of Y could be the pH of a water sample from a lake and U could be thought of as the average value of all possible measurements Y made on this lake according to a specified protocol. In this setting, U is a random variable if the lake is selected at random from a specified population of lakes. The function F_U is of interest when the objective is to estimate characteristics of the population of lakes, excluding within-lake variability.

The model that we assume for the data is as follows. The observable random variables are $Z_{ij} = U_i + \varepsilon_{ij}$, where the U_i's ($i = 1, 2, \ldots, n$) are an iid sample from the distribution F_U and the ε_{ij}'s ($i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n_i$) are iid normal random variables with mean 0 and variance σ^2. Because some knowledge of σ^2 is essential before its effect can be modeled, we assume that enough of the n_i's are larger than 1 so that an adequate estimate of σ^2 can be obtained. Once σ^2 is estimated from the Z_{ij}'s, we treat it as known and reduce the data to the within-unit averages $Y_i = \sum_j Z_{ij}/n_i$. Thus the specific problem that we address is estimation of F_U based on the observations $Y_i = U_i + \delta_i$, where U_i and δ_i are independent and δ_i is normally distributed with known variance, σ^2/n_i.

Because the distribution of each Y_i is the convolution of the distributions of U_i and δ_i, estimation of F_U has been referred to as deconvolution. In a more general setting, this problem can also be viewed as the estimation of a mixing distribution. Maritz and Lwin (1989, chap. 2) described several approaches to the estimation of a mixing distribution, which in empirical Bayes methods is referred to as a prior distribution, and the related issue of identifiability. Laird (1978) considered nonparametric maximum likelihood estimation of a mixing distribution and showed that the resulting estimates are self-consistent. A number of authors (e.g., Devroye 1989; Hesse 1995; Liu and Taylor 1989; Snyder, Miller, and Schulz 1988; Stefanski and Carroll 1990, Zhang 1990) considered using kernel density estimators for density deconvolution and gave corresponding convergence results. Density deconvolution has also been considered by Mendelsohn and Rice (1982) using B-splines and by Masry and Rice (1992), who used estimates of the derivatives of the convoluted density. Estimators for deconvoluting a distribution function in the presence of normally distributed measurement errors have been presented by Gaffy (1959) and Stefanski and Bay (1994). Scheinok (1964) gave an estimator for the analogous problem where the errors are exponentially distributed. Optimal rates of convergence in nonparametric deconvolution problems have been studied by Carroll and Hall (1988) and Fan (1991). These authors showe d that rates of convergence are very slow, especially when the error distribution is smooth.

Our approach, which was also considered by Maritz and Lwin (1989), is to model $F_U(t)$ as a mixture of a finite umber of known distributions. The mixture model provides the flexibility to approximate a wide range of distributions. I n many applications it is reasonable to assume that the distribute of U is unimodal, so we also consider imposing the additional constraint of unimodality. In both models parameters are estimated and large-sample confidence intervals are constructed using well-known likelihood theory.

In Section 2 the estimation of F_U using the EM algorithm is described. Calculation of confidence intervals for F_U is treated in Section 3. The methods developed in Sections 2 and 3 are extended in Section 4 to obtain a unimodal estimator of F_U and corresponding confidence intervals. In Section 5 the application of these estimators is illustrated with data obtained in a dietary study. The results of a simulation study, designed to indicate the performance of the estimators, are given in Section 6.

2. ESTIMATION IN THE MIXTURE MODEL

In this section we specify a mixture model for F_U and describe maximum likelihood estimation under this model. Assume that F_U can be represented as a mixture, $F_U = \sum_{k=1}^{m} p_k F_k$, where the F_k's are known distribution functions and the p_k's are unknown nonnegative constants satisfying $\sum_{k=1}^{m} p_k = 1$. We refer to the distributions F_k as

Clifford B. Cordy is Software Engineer, Accu-Fab Systems Inc., Corvallis, OR 97330. David R. Thomas is Professor, Department of Statistics, Oregon State University, Corvallis, OR 97331. This research was supported by Cooperative Agreement CR-821800 between the United States Environmental Protection Agency and Oregon State University. The authors thank the referees for their helpful comments.

Journal of the American Statistical Association
December 1997, Vol. 92, No. 440, Theory and Methods

the component distributions. Specific choices of the F_k's are described later in this section and in Section 4. Under the mixture model, the distribution of each Y_i can also be expressed as a mixture,

$$F_{Y_i} = \sum_{k=1}^{m} p_k(F_k * \phi_i), \tag{1}$$

where ϕ_i is the density for the normal distribution with mean 0 and variance σ^2/n_i, and $*$ denotes convolution. Assume that each F_k has a corresponding density function f_k. Then the log-likelihood function for the mixing proportions $\mathbf{p} = (p_1, \ldots, p_m)^T$ based on the data $\mathbf{y} = (y_1, \ldots, y_m)^T$ is given by

$$L(\mathbf{p}|\mathbf{y}) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{m} p_k(f_k * \phi_i)(y_i) \right).$$

Although it may be difficult to directly maximize the log-likelihood with respect to p, as was noted by Maritz and Lwin (1989, p. 54), the EM algorithm (Dempster, Laird, and Rubin 1977) is quite simple to apply for the estimation of mixing proportions. For the purpose of applying the EM algorithm to our problem, the complete data consist of the pairs (y_i, k_i), where k_i denotes the index of the component of the mixture (1) from which the observation y_i was made. The sufficient statistics for $\mathbf{p}$ from the complete data are the counts C_k = the number of i's for which k_i equals k. Given a prior estimate of $\mathbf{p}$, say $\mathbf{p}^{\text{old}}$, an updated estimate is obtained through the following steps:

E step: For $k = 1, \ldots, m$ calculate the conditional expectations

$$\hat{C}_k = E(C_k|\mathbf{y}, \mathbf{p}^{\text{old}}) = \sum_{i=1}^{n} \left(\frac{p_k^{\text{old}}(f_k * \phi_l)(y_i)}{\sum_{l=1}^{m} p_l^{\text{old}}(f_l * \phi_i)(y_i)} \right).$$

M step: Calculate the maximum likelihood estimate of $\mathbf{p}, \mathbf{p}^{\text{new}}$, from the estimates $\hat{C}_1, \ldots, \hat{C}_m$:

$$p_k^{\text{new}} = \hat{C}_k/n.$$

Once the estimate, say $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_m)^T$, of $\mathbf{p}$ is obtained, the corresponding estimate of $F_U(t)$ is given by

$$\hat{F}_U(t) = \sum_{k=1}^{m} \hat{p}_k F_k(t).$$

Regarding the choice of the component distributions, a minimal requirement is that the vector of mixing proportions be identifiable. This will be the case when the functions $F_1, \ldots, F_m$ are linearly independent (Teicher 1963). A simple choice for the component distributions is to take them to be normal distributions with a common variance, σ_c^2, and means that are equally spaced within the range of the observations. The choice of the number, m, of components and the value of σ_c^2 can have a large impact of the performance of the resulting estimator. The variance of $\hat{F}_U(t)$ usually increases as m increases. On the other hand, when m is too small, the mixture model may provide only a poor approximation to the true distribution, in which case the bias of $\hat{F}_U(t)$ will be unacceptably large. Thus it is usually desirable to choose the smallest value of m that gives an adequate approximation to F_U. The effect of σ_c^2 is similar but in the opposite direction. The variance of $\hat{F}_U(t)$ tends to decrease as σ_c^2 increases, but choosing σ_c^2 too large will result in increased bias. These effects are illustrated in the simulations presented in Section 6. Other choices of component distributions, such as the B splines used by Mendelsohn and Rice (1982) or the discrete point masses or uniform distributions considered by Maritz and Lwin (1989), have also bee considered. The actual shape of the component distributions has some effect but is relatively unimportant. For simplicity, we use normal component distributions, because the convolution (required in the E step) of a normal component density with the normal measurement error density is itself a normal density.

3. CONFIDENCE INTERVALS FOR $F_U(t)$

We now describe two methods for constructing a confidence interval for $F_U(t)$ at a fixed value of t. The first method is based on a normal approximation to the distribution of $\hat{F}_U(t)$ with an estimated standard error calculated using the inverse of observed information matrix. The second method relies on a chi-squared approximation to the distribution of the likelihood ratio statistic. In both methods we treat σ as known even though the value of this parameter is unknown and must be estimated. As a possible alternative, the likelihood function can be expressed in terms of the original variables Z_{ij}. Inferences can then be made with respect to the full parameter space in which σ is unknown. Although this approach may be preferable when σ is imprecisely estimated, we do not pursue this approach in this article.

The confidence interval, with nominal confidence coefficient $1 - \alpha$, based on the normal approximation, is given by

$$[\hat{F}_U(t) - z_{\alpha/2}\hat{\text{SE}}(\hat{F}_U(t)), \hat{F}_U(t) + z_{\alpha/2}\hat{\text{SE}}(\hat{F}_U(t))], \tag{2}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution and $\hat{\text{SE}}(\hat{F}_U(t))$ is an estimate of the standard error of $\hat{F}_U(t)$. The following discussion describes standard error estimation. The variance of $\hat{F}_U(t)$ is given by

$$\text{var}(\hat{F}_U(t)) = \sum_{r=1}^{m}\sum_{s=1}^{m} F_r(t)F_s(t)\text{cov}(\hat{p}_r, \hat{p}_s). \tag{3}$$

Estimates of the covariances in this expression for $r < m$ and $s < m$ are given by the entries of the matrix $\mathbf{I}^{-1}$, where $\mathbf{I}$ is the observed information matrix for the parameters $p_1, \ldots, p_{m-1}$ and has entries

$$\begin{aligned} \mathbf{I}_{rs} &= -\frac{\partial^2 L(\mathbf{p}|\mathbf{y})}{\partial p_r \partial p_s} \\ &= -\sum_{i=1}^{n} \left\{ \frac{[f_r * \phi_i(y_i) - f_m * \phi_i(y_i)] \times [f_s * \phi_i(y_i) - f_m * \phi_i(y_i)]}{[\sum_k \hat{p}_k f_k * \phi_i(y_i)]^2} \right\}. \end{aligned}$$

The estimated covariances for $r = m$ can then be obtained using the formula

$$\operatorname{cov}(\hat{p}_m, \hat{p}_s) = \operatorname{cov}\left(1 - \sum_{r<m} \hat{p}_r, \hat{p}_s\right) = -\sum_{r<m} \operatorname{cov}(\hat{p}_r, \hat{p}_s).$$

The case $s = m$ is handled similarly. The estimated standard error in (2) can be calculated by substituting these estimated covariances into (3) and taking the square root of the resulting variance estimate.

We now describe calculation of a confidence interval for $F_U(t)$ using the likelihood ratio statistic. This statistic is given by $2[L^*(\hat{F}) - L^*(F)]$, where $L^*(F)$ is the profile log-likelihood for $F = F_U(t)$ and $\hat{F}$ is the maximum likelihood estimator of $F_U(t)$. The profile log-likelihood function is defined by

$$L^*(F) = \sup\left\{L(\mathbf{p}|\mathbf{y}):\ F = \sum p_k F_k(t) = \sum p_k F_k\right\}$$

for those values of F satisfying $\min\{F_k\} \le F \le \max\{F_k\}$. (For a general discussion of the profile likelihood and its use in constructing confidence intervals see, e.g., Barndorff-Nielson and Cox 1994, sec. 3.4.) An approximate $(1-\alpha) \times 100\%$ confidence interval for $F_U(t)$ is the set of values F such that

$$2[L^*(\hat{F}) - L^*(F)] \le \chi^2_\alpha, \tag{4}$$

where χ^2_α is the $1-\alpha$ quantile of the chi-squared distribution with 1 df.

The function $L^*(F)$ is somewhat difficult to calculate, because its value is the result of a constrained maximization. Fortunately, the EM algorithm can also be used for this purpose. The E-step is the same as in the calculation of the maximum likelihood estimator. The M step requires a restricted maximization, although a much simpler one than suggested by the definition of $L^*(F)$. Explicitly, given the expected counts $\hat{C}_1, \ldots, \hat{C}_m$, the M step is to maximize

$$\sum_{k=1}^{m} \hat{C}_k \log p_k$$

subject to the constraints $\sum p_k = 1$ and $\sum p_k F_k = F$. This is the maximization of a multinomial likelihood subject to a linear constraint. The solution can be expressed in terms of a single Lagrange multiplier as

$$\hat{p}_{k,\lambda} = \frac{\hat{C}_k}{n - \lambda(F - F_k)},$$

where λ is chosen so that the two constraints given earlier are satisfied. A numerical approximation to this value of λ is easily obtained because it is the unique root of the increasing function

$$\begin{aligned} h(\lambda) &= \sum_{k=1}^{m} \frac{\hat{C}_k(F - F_k)}{n - \lambda(F - F_k)} \\ &= \left(\left(\sum_{k=1}^{m} \hat{p}_{k,\lambda}\right) - 1\right) F - \left(\left(\sum_{k=1}^{m} \hat{p}_{k,\lambda} F_k\right) - F\right) \end{aligned}$$

defined on the set of λ that satisfy $n > \lambda(F - F_k)$ for $k = 1, \ldots, m$.

4. ESTIMATION OF A UNIMODAL DISTRIBUTION

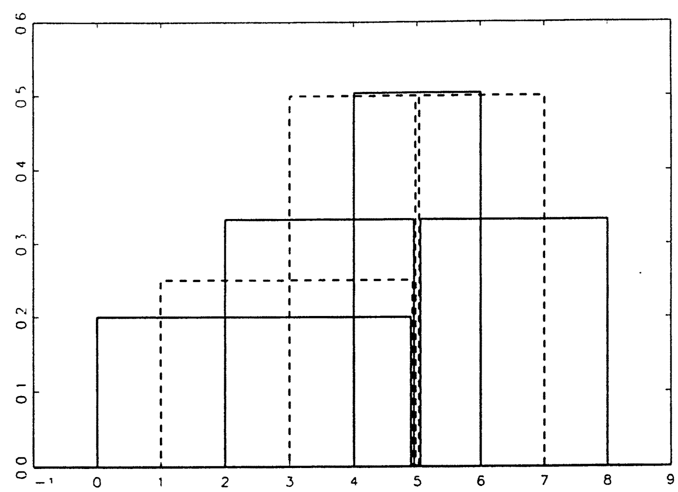
We now show how the foregoing methods can be applied to estimate a unimodal distribution (i.e., a distribution whose density is unimodal.) Our approach is to specify a family of unimodal distributions that is a finite union of subfamilies that can each be parameterized with a mixture model. Although there are a number of possible choices for such a family of unimodal distributions, in this section we focus on just one such family. At the end of this section we briefly describe an alternative family.

To specify the family of unimodal distributions, we start with equally spaced points $t_0 < t_1 < \cdots < t_{m+1}$. The subfamilies, indexed by $k = 0, \ldots, m+1$, are denoted by $\mathbb{U}(k)$. For $k = 2, \ldots, m-1$, the subfamily $\mathbb{U}(k)$ is defined to be the set of mixtures of the uniform distributions on the following m intervals: $(t_0, t_k), (t_1, t_k), \ldots, (t_{k-2}, t_k), (t_{k-1}, t_{k+1}), (t_k, t_{k+2}), \ldots, (t_k, t_{m+1})$. The densities of the components for such a mixture are illustrated in Figure 1a. The subfamily $\mathbb{U}(0)$ is the set of mixtures of the uniform distributions on the m intervals: $(t_0, t_2), (t_0, t_3), \ldots, (t_0, t_{m+1})$. For $k = 1, m$, and $m+1$, the subfamilies $\mathbb{U}(k)$ are defined in a similar manner. Because, as is shown later, each of the subfamilies $\mathbb{U}(k)$ contains only unimodal distributions the family, $\mathcal{U}$, defined to be the union of the sets $\mathbb{U}(k)$, contains only unimodal distributions.

We now verify that each distribution in the subfamily $\mathbb{U}(k)$ is unimodal. For the sake of concreteness, consider the case with $m = 7$ and $k = 5$, as in Figure 1a. The general case is handled similarly. For $1 \le i < k$, let $f_\imath$ denote the value of the uniform density function on the interval (t_{i-1}, t_k) for x in this interval, and let $p_1, \ldots, p_m$ be nonnegative constants that sum to one. Let f denote the density function for the distribution in $\mathbb{U}(k)$ with mixing proportions $p_1, \ldots, p_m$. For $x < t_0 f(x) = 0$, for $x \in (t_0, t_1) f(x) = p_1 f_1$, and for $x \in (t_1, t_2) f(x) = p_1 f_1 + p_2 f_2$. Because each p_i and $f_\imath$ is nonnegative, f is monotonically increasing from $-\infty$ to t_2. Continuing in this manner, one sees that f is monotonically increasing from $-\infty$ to t_k and monotonically decreasing from t_k to ∞. In other words, f is unimodal.

To calculate the maximum likelihood estimate (MLE) of F_U from within $\mathcal{U}$, we first calculate the MLE of F_U within each $\mathbb{U}(k)$. Because $\mathbb{U}(k)$ is parameterized as a mixture, the MLE of F_U within $\mathbb{U}(k)$ can be obtained as in Section 2. The MLE within $\mathcal{U}$ is then the estimate from the $\mathbb{U}(k)$ that yields the largest likelihood.

Due to the nature of the parameter space in the unimodal model, constructing a confidence interval for $F_U(t)$ can be more difficult than in the unconstrained case. However, if one is willing to restrict attention to the subfamily $\mathbb{U}(k)$ from which the the MLE is obtained, instead of the full family $\mathcal{U}$, then either of the methods described in Section 3 can be applied directly within $\mathbb{U}(k)$. This is similar to assuming that the mode is known to be in the interval (t_{k-1}, t_{k+1}) and making inferences under this assumption. As is shown



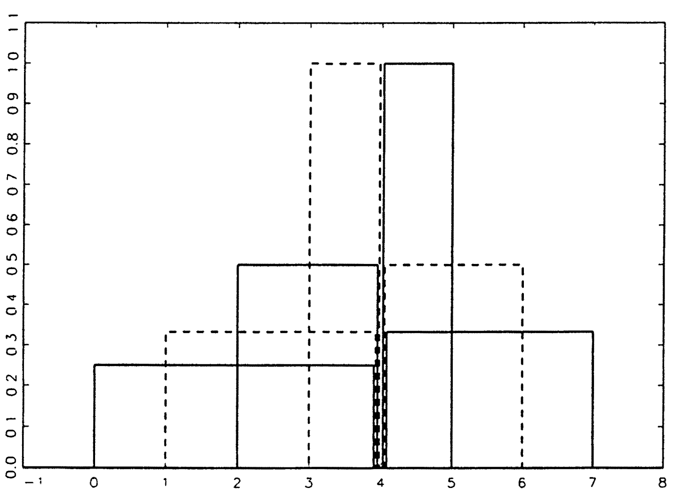


Figure 1. Density Functions for the Mixture Components for Distributions in U(k) With (a) k = 5 and (b) k = 4. In (a) m = 7 and $t_k = k$ for k = 0, . . . , 8. In (b), m = 7 and $t_k = k$ for k = 0, . . . , 7. Overlapping lines are slightly offset in this figure so that the individual densities can be distinguished.

in the simulation results in Section 6, this approach is often adequate. However, this approach may be unsatisfactory if the mode is poorly estimated. An alternative approach, not based on the restriction to the subfamily $\mathbb{U}(k)$, is to use the profile log-likelihood

$$L^*(F) = \max_{1 \leq k \leq m} L_k^*(F),$$

where $L_k^*(F)$ is the profile log-likelihood within the subfamily $\mathbb{U}(k)$. An approximate confidence interval for F can be calculated using the chi-squared approximation (4). The function $L_k^*(F)$ can be calculated as described in Section 3. Further investigation is needed to assess the adequacy of the chi-squared approximation in this setting.

We close this section by mentioning a slight variation of the foregoing model for a unimodal distribution. In this model we also start with equally spaced points $t_0 < t_1 < \cdots < t_m$. Each subfamily $\mathbb{U}^*(k)$ is defined to be the set of mixtures of the uniform distributions on the following m intervals: (t_0, t_k), $(t_1, t_k), \ldots, (t_{k-1}, t_k)$, (t_k, t_{k+1}), $(t_k, t_{k+2}), \ldots, (t_k, t_m)$. The densities of the components for

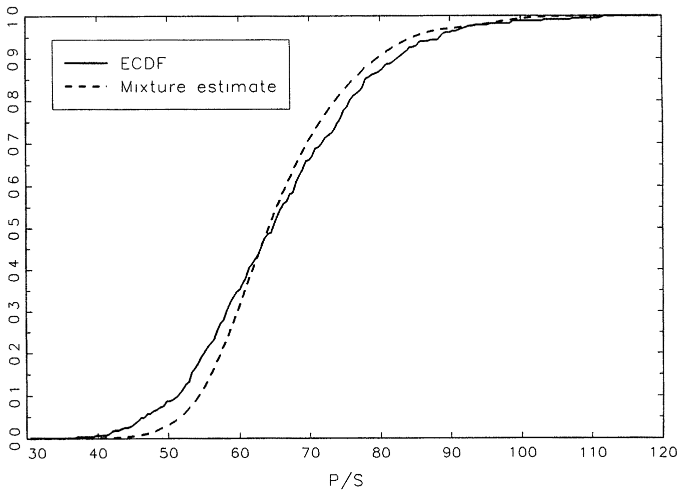


Figure 2. ECDF (——) and Deconvolution Estimate (- - -) of the Distribution Function of P/S.

such a mixture are illustrated in Figure 1b. This model, although somewhat simpler than the other, results in a unimodal estimator whose variance is larger than the unimodal estimator described earlier.

5. EXAMPLE

In a pilot study, reported by Clayton (1992), the ratio of polyunsaturated to saturated fat intake (P/S) was measured for 336 males in a 1-week full-weighted dietary survey. We follow Clayton's lead and treat the measured values of P/S for the ith individual as normally distributed with mean equal to the true value, U_i, of P/S and constant measurement error variance. Because of the men completed two such surveys separated by approximately 6 months, the variance is estimated from these repeated measurements. In the subsequent analysis, the data consists of the average value of the two measurements made on each of these 76 individuals together with the single measurements made on the others.

Figure 2 shows the empirical cumulative distribution function (ecdf) and our mixture estimator of the distribu-

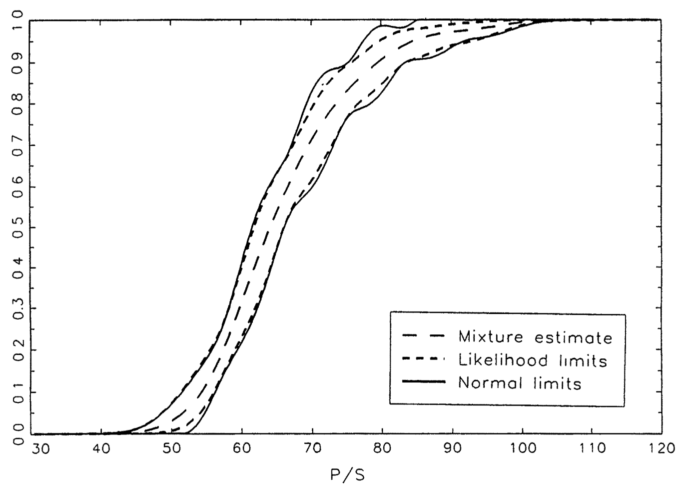


Figure 3. Nominal 95% Confidence Limits for $F_U(t)$. ——, normal limits; - - -, likelihood limits; – – –, the mixture estimate.

Table 1. Empirical Bias and Standard Errors for Deconvolution Estimators

	$F_U(t)$				
Estimator	*.10*	*.30*	*.50*	*.70*	*.90*
			Normal distribution		
ECDF	.024 (.033)	.019 (.047)	−.003 (.049)	−.022 (.047)	−.027 (.034)
Mix; U,7,L	−.003 (.043)	.000 (.063)	−.003 (.069)	−.004 (.063)	.000 (.044)
Mix; N,7,L	−.003 (.044)	.000 (.064)	−.002 (.071)	−.005 (.065)	.000 (.045)
Mix; N,7,M	−.003 (.036)	.000 (.056)	−.003 (.062)	−.005 (.057)	.000 (.038)
Mix; N,7,H	−.001 (.031)	−.002 (.046)	−.002 (.052)	−.003 (.048)	−.002 (.031)
Mix; N,5,M	−.002 (.032)	−.003 (.050)	−.003 (.056)	−.001 (.053)	−.000 (.033)
Mix; N,9,M	−.002 (.040)	.000 (.061)	−.003 (.068)	−.003 (.062)	−.001 (.041)
Unim1,7	−.003 (.036)	−.001 (.052)	−.003 (.057)	−.004 (.053)	.001 (.038)
Unim2,7	−.004 (.037)	−.002 (.055)	−.003 (.061)	−.002 (.056)	.001 (.039)
			Extreme value distribution		
ECDF	.019 (.031)	.027 (.046)	.012 (.049)	−.018 (.046)	−.048 (.035)
Mix; U,7,L	.001 (.037)	−.003 (.060)	−.003 (.064)	.001 (.062)	.006 (.047)
Mix; N,7,L	.001 (.038)	−.003 (.061)	−.003 (.068)	.000 (.068)	.007 (.045)
Mix; N,7,M	.000 (.034)	−.003 (.053)	.000 (.059)	.007 (.056)	.002 (.036)
Mix; N,7,H	.000 (.029)	.017 (.048)	.027 (.050)	.011 (.045)	−.025 (.035)
Mix; N,5,M	−.001 (.031)	.005 (.051)	.020 (.056)	.016 (.050)	−.014 (.038)
Mix; N,9,M	.000 (.036)	−.001 (.058)	−.006 (.065)	.000 (.060)	.004 (.038)
Unim1; 7	−.001 (.032)	.001 (.051)	.006 (.054)	.003 (.049)	−.003 (.040)
Unim2; 7	.000 (.032)	−.004 (.054)	−.002 (.060)	.003 (.059)	.005 (.046)

tion function for U. The component distributions used in the mixture model were seven normal distributions with equally spaced means. The variance of each component was chosen to correspond to the medium level of σ_c^2 used in the simulations, as described in the next section. The deconvolution estimator of the underlying distribution is more skewed than the ecdf. This is not surprizing, because the convolution of a skewed distribution with a symmetric error distribution will not be as skewed as the original distribution.

Both types of pointwise 95% confidence intervals, described in Section 3, are shown in Figure 3. Note that the two types of intervals agree fairly closely except for the bulges in the intervals based on the normal approximation. These bulges are also present in the intervals based on the profile likelihood, but are much less pronounced. The simulation results given in the next section indicate that in terms of coverage properties, the intervals based on the profile likelihood may be preferable.

6. SIMULATION RESULTS

We now present the results of two simulations to indicate the performance of the deconvolution estimators. These two simulations were part of a larger simulation study designed to assess the effects of various factors on the performance of deconvolution estimators. The factors considered include the shape of F_U, the variance of the measurement error, sample size, and the number and shape of the component distributions in the mixture model. In the simulations presented here, the sample size is held constant at $n = 100$ and the measurement error variance is held constant at $\frac{1}{2}$ the variance of U. In the first simulation, the random variable U had an extreme value distribution; in the second, it had a standard normal distribution. The observations used in the simulations were generated using the Gauss random number generators RNDU and RNDN. The results for both simulations are based on 1,000 simulated samples.

The results from the simulations are summarized in Table 1. The estimators calculated in the simulations were the ecdf and the mixture and unimodal deconvolution estimators. To investigate their effects, the shape of the component distributions (denoted by N and U for normal and uniform), the number of components ($m = 5, 7, 9$), and σ_c^2, the variance of the components [at three levels: low (L), medium (M), and high (H)], were varied in the specification of the mixture model. In each realization, the supports of the uniform component distributions were taken to be m nonoverlapping intervals equally dividing the interval be-

Table 2. Empirical Mean of the Standard Error Estimator

	$F_U(t)$				
Estimator	*.10*	*.30*	*.50*	*.70*	*.90*
			Normal distribution		
Mix; N,7,M	.046 (.036)	.071 (.056)	.077 (.062)	.070 (.057)	.047 (.038)
Unim1,7	.048 (.036)	.062 (.052)	.065 (.057)	.062 (.053)	.047 (.038)
			Extreme value distribution		
Mix; N,7,M	.050 (.034)	.071 (.053)	.073 (.059)	.064 (.056)	.045 (.036)
Unim1; 7	.055 (.032)	.065 (.051)	.061 (.054)	.055 (.049)	.043 (.040)

NOTE: The observed standard error of the deconcolution estimator is given in parentheses.

Table 3. Empirical Coverage Probabilities for Nominal 95% Confidence Intervals Using the Inverse Information and Profile Likelihood Methods

Estimator/ method	$F_U(t)$ .10	.30	.50	.70	.90
			Normal distribution		
Mix; N,7,M/					
Information	.977	.971	.973	.983	.978
Likelihood	.961	.952	.954	.059	.965
Unim1; 7/					
Information	.981	.958	.965	.967	.972
Likelihood	.951	.934	.928	.930	.945
			Extreme value distribution		
Mix; N,7,M/					
Information	.993	.984	.975	.958	.986
Likelihood	.970	.959	.953	.948	.968
Unim1; 7/					
Information	.992	.970	.967	.961	.944
Likelihood	.967	.929	.929	.929	.885

tween the smallest to the largest observed value of Y. So the normal components and the uniform components could be directly compared, we used normal components with the same means and variance as the uniform components. This specifies the low level of σ_c^2. The medium level of σ_c^2 is 4 times as large as the low level, and the high level of σ_c^2 is 12 times as large as the low level. At the high level, σ_c is equal to the distance between the component means. The points $t_0, \ldots, t_{m+1}$ used for unimodal estimation were chosen so that t_0 and t_{m+1} agree with the smallest to the largest observed values of Y. Estimators based on the two unimodal parameterizations (denoted by Unim1 and Unim2) described in Section 4 were also calculated.

Most of the deconvolution estimators considered effectively reduced bias relative to the ecdf. The exceptions were at the low level of $m, m = 5$, and the high level of σ_c^2 when U follows the extreme value distribution. It appears that the price of bias reduction is an increase in variance. This increase in variance depends on the choice of m and σ_c^2. Increasing m and/or decreasing σ_c^2 increases the variance of the resulting deconvolution estimator. On the other hand, choosing m too small or σ_c^2 too large, as in the two aforementioned cases, can result in a deconvolution estimator with a substantial bias. Thus a reasonable strategy would be to try to choose the smallest value of m and the largest value of σ_c^2 that yield an acceptably small bias. In the cases presented here, we feel that $m = 7$ and σ_c^2 at the medium level is a reasonable compromise. The shape of the components appears to have little effect on the bias of the mixture estimator, although using the uniform components results in a slightly smaller variance than when the normal components are used. However, this difference is relatively minor when compared to the effect of varying either m or σ_c^2. The unimodal estimators have a similar level of bias, but the estimator Unim1 has a smaller variance.

The performance of the standard error estimator and the confidence interval procedures described in Section 3 were considered for two of the estimators. These estimators were the mixture estimator using normal components with the medium level of σ_c^2 and the unimodal estimator, Unim1, both with $m = 7$. The results are given in Tables 2 and 3. For the unimodal estimator, the standard error estimates and the confidence intervals were calculated within the restricted model $\mathbb{U}(k)$, as described in Section 4, not in the full unimodal model. The standard error estimator tends to overestimate the true standard error in all cases. Thus it may be worth investigating alternative standard error estimators in this setting. Coverage probabilities are reasonably close to nominal levels in all cases considered in Table 3. More specifically, coverage probabilities for the intervals using the inverse information (normal approximation) tend to be somewhat larger than the nominal level. For the mixture estimator, the coverage probabilities for the intervals based on the profile likelihood are quite close to the nominal level. For the unimodal estimator, the coverage probabilities for this method are slightly smaller than the nominal level. Using the full likelihood for the unimodal model, as suggested in Section 4, may give coverage probabilities closer to the nominal level.

7. SUMMARY

The ordinary ecdf is a biased estimator of a distribution function when only a contaminated version of the variable of interest can be observed. Much of this bias can be removed through the use of a deconvolution estimator as is developed in this article.

The mixture model that we have considered will give fairly close approximations to a wide range of possible distributions. Although such parameterizations are reminiscent of a nonparametric approach, when compared to a completely nonparametric formulation they have the advantages provided by maximum likelihood theory. In particular, the resulting MLEs are more stable than nonparametric estimators, and estimated standard errors or confidence intervals can be calculated using likelihood theory. We also showed that the same theory can be applied to obtain a unimodal deconvolution estimator.

Our simulation results indicate that the deconvolution estimators considered are effective at reducing bias relative to the ecdf. The variance of the resulting estimator will be somewhat larger than that of the ecdf, but the amount of additional variance can be controlled through a careful choice of parameterization. The simulations also indicate that the coverage properties of confidence intervals are satisfactory.

[Received July 1995. Revised February 1997.]

REFERENCES

Barndorff-Nielson, O. E., and Cox, D. R. (1994), *Inference and Asymptotics*, London: Chapman and Hall.

Carroll, R., and Hall, P. (1988), "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83, 1184–1186.

Clayton, D. G. (1992), "Models for the Analysis of Cohort and Case-Control Studies With Inaccurately Measured Exposures," in *Statistical Models for Longitudinal Studies of Health*, eds. J. H. Dwyer, M. Feinleib, P. Lippert, and H. Hoffmeister, New York: Oxford University Press, pp. 301–331.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal* *of the Royal Statistical Society* Ser. B, 39, 1–38.

Devroye, L. (1989), "Consistent Deconvolution in Density Estimation," *The Canadian Journal of Statistics*, 17, 235–239.

Dharmadhikari, S., and Joav-Dev, K. (1988), *Unimodality, Convexity, and Applications*, San Diego, Academic Press.

Fan, J. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19, 1257–1272.

Gaffy, W. (1959), "A Consistent Estimator of a Component of a Convolution," *Annals of Mathematical Statistics*, 30, 198–205.

Hesse, C. H. (1995), "Deconvolving a Density from Partially Contaminated Observations," *Journal of Multivariate Analysis*, 55, 246–260.

Laird, N. (1978), "Nonparametric Maximum Likelihood Estimation of a Mixing Distribution," *Journal of the American Statistical Association*, 73, 805–811.

Liu, M. C., and Taylor, R. L. (1989), "A Consistent Nonparametric Density Estimator for the Deconvolution Problem," *The Canadian Journal of Statistics*, 17, 427–438.

Maritz, J. S. (1968), "On the Smooth Empirical Bayes Approach to Testing of Hypotheses and the Compound Decision Problem," *Biometrika*, 55, 83–100.

Maritz, J. S., and Lwin, T. (1989), *Empirical Bayes Methods*, New York: Chapman and Hall.

Masry, E., and J. A. Rice (1992), "Gaussian Deconvolution via Differentiation," *The Canadian Journal of Statistics*, 20, 9–21.

Mendelsohn, J., and Rice, J. (1982), "Deconvolution of Microflourometric Histograms With B-Splines," *Journal of the American Statistical Association*, 77, 748–753.

Scheinok, P. (1964), "Estimation of a Component of a Convolution, When the Other Component is of the Exponential Type," *Technometrics*, 6, 222–224.

Snyder, D. L., Miller, M. I., and Schulz, T. J. (1988), "Constrained Probability-Density Estimation From Noisy Data," in *Proceedings of the 1988 Conference on Information Science and Systems*, pp. 170–172.

Stefanski, L., and Bay, J. (1994), "Parametric Jackknife Deconvolution of Finite Population CDF Estimators," Institute of Statistics Mimeograph Series No. 2262, North Carolina State University.

Stefanski, L., and Carroll, R. (1990), "Deconvoluting Kernel Density Estimators," *Statistics*, 21, 169–184.

Teicher, H. (1963), "Identifiability of Finite Mixtures," *Annals of Mathematical Statistics*, 34, 1265–1269.

Zhang, C. (1990), "Fourier Methods for Estimating Mixing Densities and Distributions," *The Annals of Statistics*, 18, 806–831.